Model Archive Summary for Suspended-Sediment Concentration at U.S. Geological Survey Station 11304810 San Joaquin River below Garwood Bridge at Stockton, California

This model archive summary details the suspended-sediment concentration (SSC) model developed to compute 15-minute SSC beginning October 1, 2010. This is the first suspended-sediment model developed for the site. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memorandum 2016.07/2016.10 (USGS, 2016) and USGS Techniques and Methods, book 3 section C, chapter 4 (Rasmussen and others, 2009). This summary and model archive are in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (USGS 2014).

Site and Model Information

Site number: 11304810
Site name: San Joaquin River below Garwood Bridge at Stockton, CA (SJG)
Location: Latitude 37°56'08", Longitude 121°19'45"referenced to North American Datum of 1927, San Joaquin County, CA, Hydrologic Unit 18040003.
Equipment: A YSI 6-series sonde began logging turbidity with a model 6136 sensor on December 1, 2009 and was removed from the station on December 29, 2014.

Model number: 11304810.SSC.WY2011.1
Model calibration data period: November 23, 2010 – December 19, 2014
Model application date: October 1, 2010 – December 29, 2014
Computed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)
Reviewed by: Anna Conlen, USGS, Sacramento, CA (aconlen@usgs.gov)

Physical Sampling Details and Sediment Data

All sediment data were collected using U.S. Geological Survey (USGS) protocols (USGS, 2006) and are stored in the National Water Information System (NWIS) database https://waterdata.usgs.gov/nwis. Discrete, boat-based sample collection for SSC monitoring ideally occurs between 6-12 times per year. Sample collection spans the range of conditions and targets storm events during winter and spring flows as well as summer low flow conditions. Sample collection at SJG varied year to year while the YSI 6-series sonde was deployed, with an average of 6 samples collected per water year.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999) and USGS (2006). Sediment samples represent the discharge-weighted concentrations of the stream cross section. The Equal Discharge Increment (EDI) method was used to determine the locations of five sampling verticals along the transect where discharge weighted suspended-sediment samples were collected. Each sampling vertical is located at the centroid of increments representing 20% of the total flow (5 verticals). Due to the tidal nature of the site, the EDI method was used to collect discharge-weighted samples to represent the average cross section because velocities are not always isokinetic (based on Table 4-5 from TWRI09A4, USGS 2006). A boat-based discharge measurement was collected immediately before sampling to determine the location of each vertical.

Technicians collected samples using either a FISP US D-74 or D-96 depth-integrated suspended-sediment sampler. Cross-section locations varied slightly but were generally downstream of the bridge, either at the gage or at the standard cross section (approximately 90 ft downstream of the gage). The channel cross-section is roughly 23 feet deep in the thalweg with a mean depth of approximately 13 feet. Sampling depths ranged from 8-20 feet depending on the tide and season. Station velocities ranged from -1.7 to +3.4 ft/sec. Sediment at this station is mostly fines (93% on average from sand/fine analysis) and potential sampling bias due to non-isokinetic sampling is considered minimal because the presence of sand is rare (90% of the time, the sand fraction is less than 2%). The sand concentration was observed around 30% following the peak high flows in the San Joaquin River during 2011.

Samples were analyzed by the USGS Sediment Laboratory in Santa Cruz, California. All samples were analyzed for sediment concentration (mg/L) by the filtration method and most samples are also analyzed for the percentage of fines (< 0.062 mm). The sand/fine break analysis can be used to identify dataset variability and potential outliers and was used to determine that sediment at this station is composed of mostly fines. Though a few of the earlier samples were composited before analysis, most EDI verticals were analyzed individually by the lab. This method of individual analysis is for quality control purposes because of rapidly changing, tidal conditions. The set average SSC of the five verticals was computed and used in the calibration model. In rare occasions when the SSC at a vertical was deemed an outlier, a manual average was computed from fewer than 5 verticals and occurred on 12/6/2013 with notes applied to the database.

All sediment data were reviewed and marked as approved in the USGS NWIS Water-Quality System database (QWDATA) and made publicly available before being included in the calibration model. Publicly available field/lab sediment data can be found at: https://waterdata.usgs.gov/nwis/uv?site_no=11304810.

Surrogate Data

Continuous 15-minute turbidity data and discharge data were collected and computed by the USGS California Water Science Center and evaluated as possible explanatory variables for SSC. Turbidity data were measured using a YSI 6-series sonde and reported in Formazin Nephelometric Turbidity Units (FNU). Turbidity data began logging on December 4, 2009 and the sonde was removed from the station on December 29, 2014 @ 14:00. All surrogate turbidity data were computed, reviewed, and approved before using in the sediment calibration model per USGS guidelines (Wagner and others 2006). Discharge data were collected, computed, reviewed, and approved by the USGS California Water Science Center and retrieved from NWIS-TS. Methods to compute discharge follow Levesque and Oberg (2012). The 15-minute timeseries data are located at https://waterdata.usgs.gov/nwis/uv?site_no=11304810.

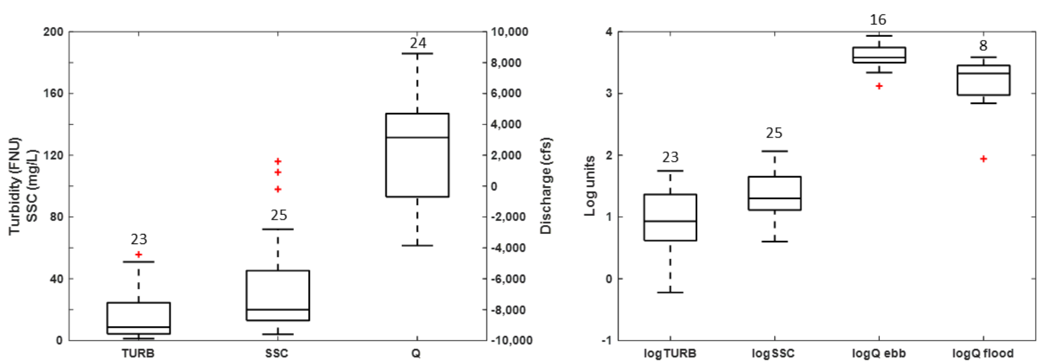
Model Calibration Dataset

The approved time-series turbidity data spanning the dates of the sediment constituent dataset were retrieved from NWIS-TS (Rasmussen and others 2009). The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair the surrogate data with the discrete sediment data (Domanski and others 2015). Turbidity and discharge values were paired with each sediment sample observation from a matching max +/- of 15 minutes. The SAID manual is found at https://pubs.er.usgs.gov/publication/ofr20151177. Of the 25 cross-sectional average sediment samples collected during the deployment, two did not have corresponding turbidity values due to deletions during records computation so were not included in the calibration dataset, leaving a total of 23 observations in the dataset. Grab and point samples were collected at this station, however because they could not be properly adjusted to be representative of the cross section per USGS guidelines, they were not included in the calibration dataset.

Also, note there was a gap in the discharge record during the sample on 6/4/2014. Summary statistics and the complete model-calibration dataset are provided in the following sections.

Regression Model Development

Multiple models were evaluated including simple linear regression (SLR) and multiple linear regression (MLR). The most common estimation technique is SLR, but MLR is an alternate tool for computing SSCs when the SLR *MSPE* statistic is larger than 20 percent (Rasmussen and others, 2009). The calibration dataset is composed of 23 concurrent turbidity, SSC, and discharge measurements. Boxplots are shown below and note that due to negative tidal discharge values during the flood tide, ebb and flood values are shown separately with the absolute values shown during flood tides. USGS (2016) *recommends* a minimum of 36 paired observations, however the turbidity sensor was replaced with a different model at the end of 2014 requiring a separate calibration.



Model diagnostics and plots for model review were output using a combination of Matlab, SAID, and the R environment (R Core Team, 2018). The regression methods used are described in Helsel and Hirsch (2002). Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate the models. The best model was chosen based on residual plots, model standard error, R^2, significance tests (p-values), correlation of explanatory variables, variance inflation factor (VIF), and PRESS (prediction error sum of squares) statistics. Values for the statistics and metrics were computed for various models and are included below along with all relevant sample data and more in-depth statistical information.

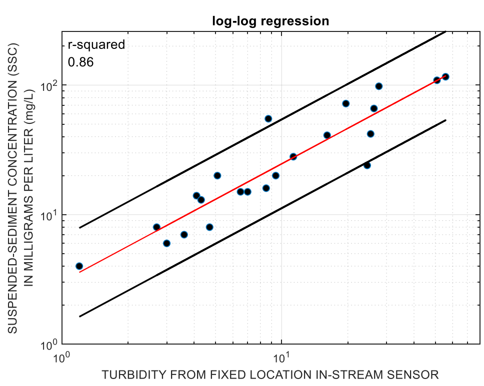
A variety of models were evaluated: Model 1) linear model with one explanatory variable (turbidity), Model 2) $\log_{10}$ transformed model with one explanatory variable (turbidity), Model 3) repeated medians method (Helsel and Hirsh, 2002) using one explanatory variable (turbidity), Model 4) linear model with two explanatory variables (turbidity and discharge), and Model 5) $\log_{10}$ transformed model with two explanatory variables (turbidity and discharge). Diagnostic statistics are summarized below for the five models evaluated. Discharge was not considered further as a second surrogate (in addition to turbidity) because the log model(s) had the best normality but the lowest PRESS statistic was between the log-transformed regressions was from the SLR model. The ability to accurately include negative discharges during flood tide is not possible as negative discharge alone was not a significant predictor of SSC. The site is variable; while sediment can be transported from both upstream and downstream, the higher concentrations coincide with fluvial events, but turbidity and discharge are still not well correlated with each other as sediment concentration is dependent on upstream watershed conditions and there is also hysteresis. Thus, discharge was not considered further.

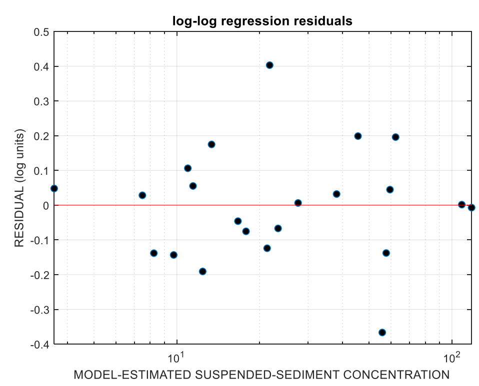
Flagged observations from the SAID outlier test criteria were evaluated. Standardized residuals from the models were inspected for values greater than 3 or less than negative 3. Values outside of the 3 to – 3 range are considered potential extreme outliers. The standardized residuals were reviewed from the SAID output reports to determine if any samples should be removed from the model.

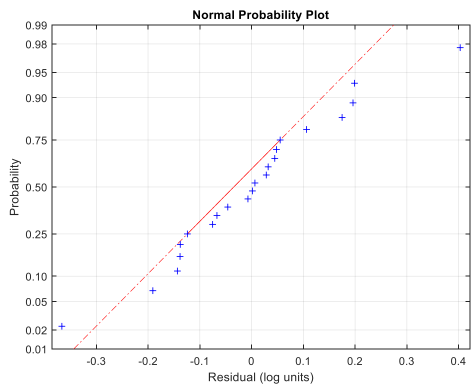
The sample collected on 2/19/2014 was flagged in SAID and it was determined that the turbidity was extremely low for the resulting SSC; and the SSC was unexpectedly low for the sand concentration. This sample appeared to be an outlier when plotted and when it was removed from the dataset, the diagnostic statistics improved. Thus, because of potential concern regarding sample results, it was removed from the final dataset for a total of n = 22 samples.

No.	R^2	R^2_a	RMSE	PRESS	MSPE	N	(type)
Model 1	0.82	0.81	15.1	5256	41.9	22	linear
Model 2	**0.86**	**0.85**	**0.2**	**0.6**	**39.1**	**22**	**log**
Model 3	0.81	0.81	15.5	5441	43.0	22	repeated median
Model 4	0.88	0.87	13.0	4396	36.1	22	multi-linear
Model 5	0.89	0.88	0.16	1.61	36.7	15	multi-log

Of the SLR models, the $\log_{10}$-transformed regression model had the lowest error statistics, the highest R^2, and the residual plots indicated a more homoscedastic pattern (constant variance) and a more normal distribution compared to the linear model(s).







Model Summary

The final regression model for suspended-sediment concentration at site number 11304810 is a simple log_{10}-transformed regression model based on 22 measurements of cross-sectional SSC samples and in situ turbidity values collected over approximately four years from November 23, 2010 to December 19, 2014. The simple linear regression model is shown below with basic model information, regression coefficients, correlation, summary statistics, and Duan's bias correction factor (Duan, 1983):

Linear Regression Model	Coefficient of Determination (R^2)
$log_{10}SSC = 0.482 + 0.91 * log_{10}Turb$	0.86

where

SSC = suspended-sediment concentration, in milligrams per liter; and

Turb = turbidity, in formazin nephelometric units, measured with a YSI model 6136.

Because SSC was transformed during regression model development, the computed prediction may be biased and needs to be multiplied by a non-parametric smearing bias correction factor (BCF) which is shown below.

Model	Start date	End date	Linear Regression Model	*BCF*
1	10/1/2010	12/29/2014	$SSC = 10^{0.482} \times Turb^{0.91} \times BCF$	1.07

The log_{10}-transformed SLR model can be retransformed and corrected for bias resulting in the following equation:

$$SSC = 3.249Turb^{0.91}$$

Parameter	Minimum	Maximum
Turbidity (FNU) entire record	0	176
Computed SSC (mg/L)	0	*358/128

The extrapolation value is shown with an asterisk above. Extrapolation is defined as computation beyond the range of the model calibration dataset may be used to extrapolate no more than 10 percent outside the range of the sample data used to fit the model. The original maximum computed SSC was 358 mg/L. The portion of time-series data beyond the extrapolation limit is less than 1%. Following USGS guidelines, the extrapolated, maximum computed SSC for this model is limited to 128 mg/L.

Suspended-Sediment Concentration Record

The complete SSC record is computed using this regression model and can be found at https://nrtwq.usgs.gov/explore/dyplot?site_no=11304810 as well as the links to all the stations in the sediment network at http://nrtwq.usgs.gov/ca.

Model

logSSC = + 0.91 * logTURB + 0.482

Variable Summary Statistics

	logSSC	SSC	logTURB	TURB
Minimum	0.602	4.0	0.0792	1.2
1st Quartile	1.110	13.0	0.6330	4.3
Median	1.300	20.0	0.9340	8.6
Mean	1.360	36.2	0.9700	14.8
3rd Quartile	1.740	55.0	1.3900	24.5
Maximum	2.060	116.0	1.7500	55.7

Basic Model Statistics

Number of Observations	22
Standard error (RMSE)	0.166
Average Model standard percentage error (MSPE)	39.1
Coefficient of determination (R^2)	0.859
Adjusted Coefficient of Determination (Adj. R^2)	0.852
Bias Correction Factor (BCF)	1.07

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.482	0.0874	5.51	2.14e-05
logTURB	0.910	0.0825	11.00	5.83e-10

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.915
E.vars	-0.915	1.000

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.273	0.193	0.603

Flagged Observations

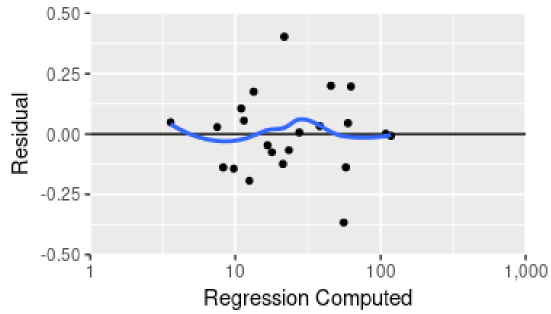
Date	Time	logSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
5/16/2012	9:26	1.74	1.34	0.403	2.49	2.92	0.0457	0.148	0.639
12/19/2014	11:32	1.38	1.75	-0.366	-2.32	-2.64	0.0890	0.262	-0.825

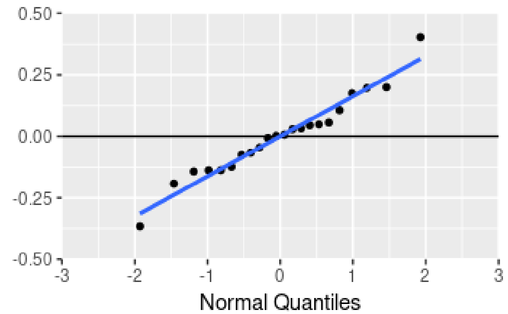
Plots of $\log_{10}$SSC and explanatory variables and residual diagnostic plots

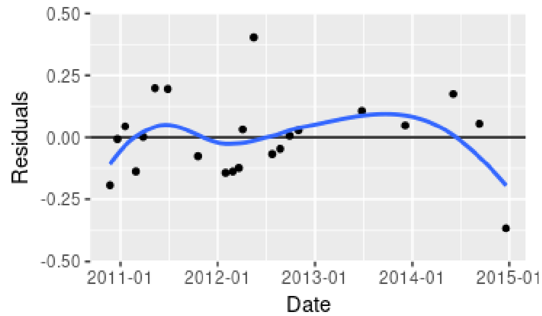
The following plots were generated online using a specialized R-Script developed by Patrick Eslick of the KSWSC and is located at the following address:

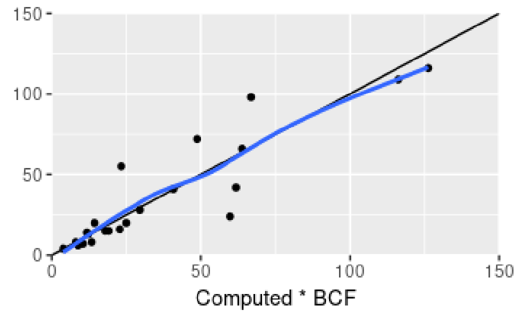
https://patrickeslick.github.io/ModelArchiveSummary/

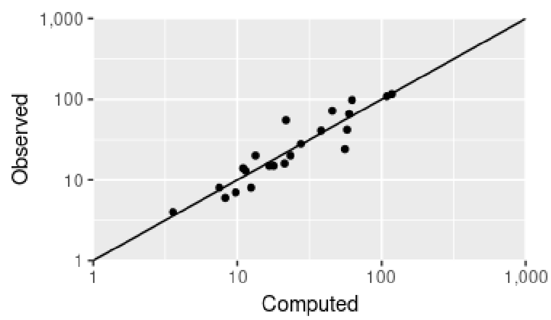
Statistical Plots

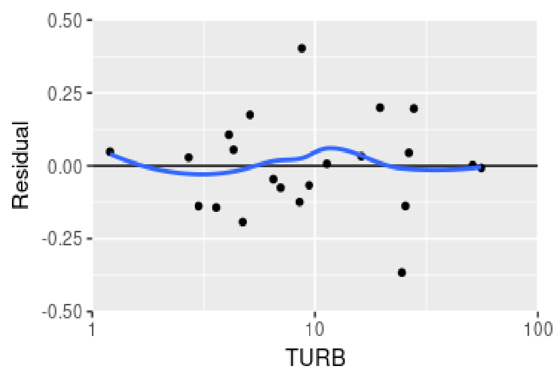


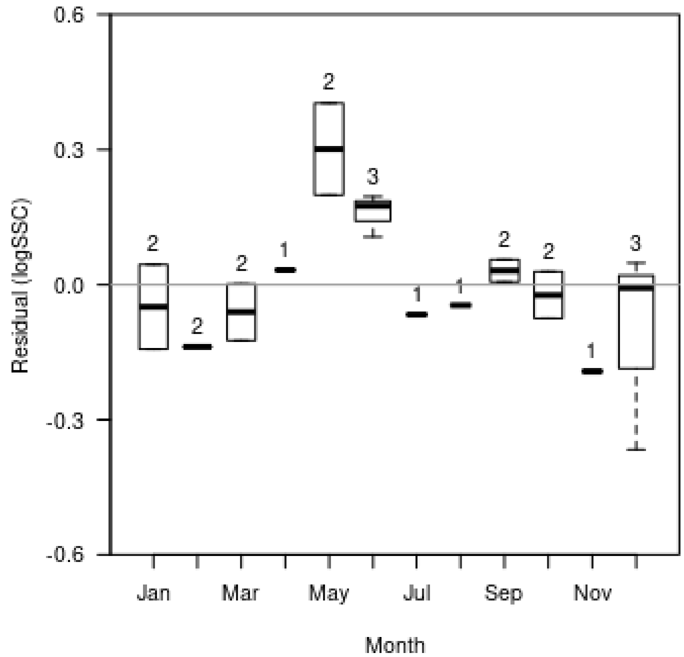


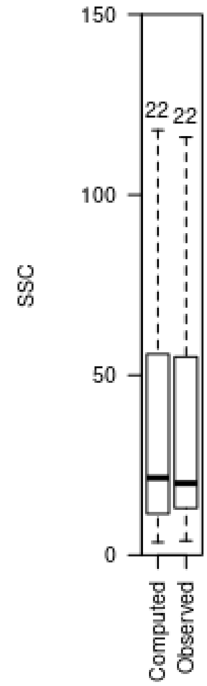


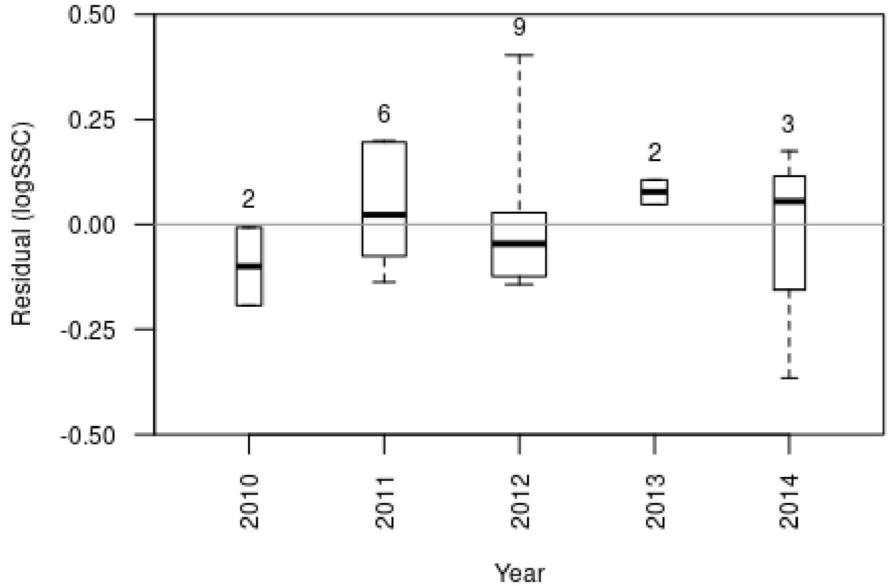




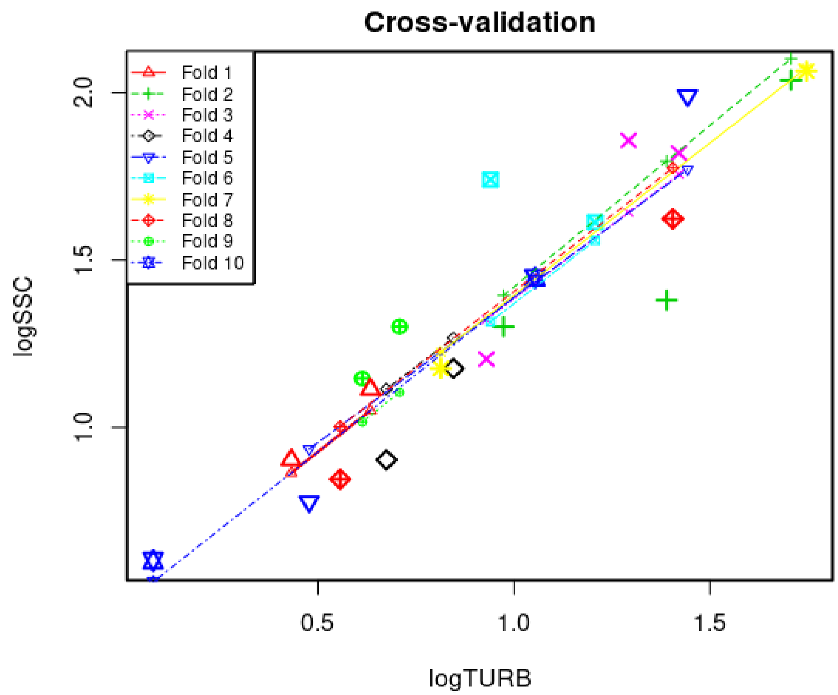



0.6
0.3
0.0
-0.3
-0.6
Residual (logSSC)
Jan
Mar
May
Jul
Sep
Nov
Month

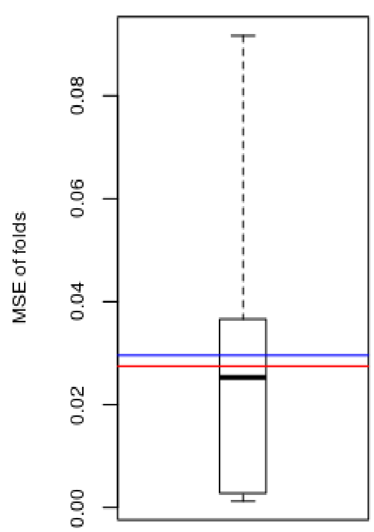

150
100
50
0
SSC
Computed
Observed


0.50
0.25
0.00
-0.25
-0.50
Residual (logSSC)
2010
2011
2012
2013
2014
Year

The graph below shows a k-fold cross validation with k=10 and the large points represent observations that were left out of each fold and are identified by the color and shape.




```
                Minimum MSE of folds:  0.00122
                   Mean MSE of folds:  0.02960
                 Median MSE of folds:  0.02530
                Maximum MSE of folds:  0.09170
 (Mean MSE of folds) / (Model MSE):  1.08000
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Calibration Dataset

Observation Number	DateTime	logSSC	logTURB	SSC	TURB	Computed logSSC	Computed SSC	Residual	Normal Quantile	Censored Values
1	11/23/2010 13:45	0.90	0.67	8	4.7	1.10	13.3513	-0.193	-1.461	--
2	12/21/2010 10:45	2.06	1.75	116	55.7	2.07	126.282	-0.007	-0.170	--
3	1/19/2011 10:05	1.82	1.42	66	26.3	1.78	63.8115	0.044	0.406	--
4	2/28/2011 12:15	1.62	1.41	42	25.4	1.76	61.7791	-0.138	-0.816	--
5	3/28/2011 12:02	2.04	1.71	109	50.9	2.04	116.206	0.002	-0.056	--
6	5/11/2011 14:50	1.86	1.29	72	19.6	1.66	48.7495	0.199	1.461	--
7	6/28/2011 10:20	1.99	1.44	98	27.7	1.80	66.827	0.196	1.190	--
8	10/19/2011 11:30	1.18	0.85	15	7.0	1.25	19.101	-0.075	-0.532	--
9	1/31/2012 10:24	0.85	0.56	7	3.6	0.99	10.426	-0.143	-1.190	--
10	2/27/2012 11:35	0.78	0.48	6	3.0	0.92	8.83134	-0.138	-0.986	--
11	3/21/2012 10:09	1.20	0.93	16	8.5	1.33	22.7944	-0.124	-0.667	--
12	4/4/2012 10:25	1.61	1.21	41	16.1	1.58	40.7753	0.032	0.286	--
13	5/16/2012 9:26	1.74	0.94	55	8.7	1.34	23.2822	0.403	1.926	--
14	7/24/2012 10:03	1.30	0.97	20	9.4	1.37	24.9818	-0.067	-0.406	--
15	8/23/2012 9:35	1.18	0.81	15	6.5	1.22	17.8547	-0.046	-0.286	--
16	9/27/2012 9:37	1.45	1.05	28	11.3	1.44	29.5403	0.007	0.056	--
17	10/31/2012 8:53	0.90	0.43	8	2.7	0.87	8.02354	0.029	0.170	--
18	6/26/2013 10:58	1.15	0.61	14	4.1	1.04	11.7366	0.106	0.816	--
19	12/6/2013 10:09	0.60	0.08	4	1.2	0.55	3.83456	0.048	0.532	--
20	6/4/2014 7:57	1.30	0.71	20	5.1	1.13	14.3167	0.175	0.986	--
21	9/10/2014 8:52	1.11	0.63	13	4.3	1.06	12.2567	0.055	0.667	--
22	12/19/2014 11:32	1.38	1.39	24	24.5	1.75	59.7601	-0.366	-1.926	--

Definitions

SSC: Suspended-sediment concentration (SSC) in mg/L (parameter code 80154)

Turb: Turbidity in FNU (parameter code 63680)

MAS App Version 1.0

References

Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., https://pubs.usgs.gov/of/2015/1177/ofr20151177.pdf.

Duan, Naihua. 1983. Smearing estimate – A nonparametric retransformation method: Journal of the American Statistical Association. Volume 78-383. 605-610 p.

Edwards TK and Glysson GD. 1999. Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 3, Chap. C2. 89 p. Available from: https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.

Levesque, V.A., and Oberg, K.A., 2012, Computing discharge using the index velocity method: U.S. Geological Survey Techniques and Methods 3-A23, 148 p. (Also available at http://pubs.usgs.gov/tm/3a23/.)

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available from: https://www.R-project.org/.

Rasmussen P, Gray JR, Glysson GD, Ziegler AC. 2009. Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data. Book 3 Applications of Hydraulics, Section C. 52 p. Available from: https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

[USGS] U.S. Geological Survey. 2006. National field manual for the collection of water quality data: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 9, Chapter A4. Available from: https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf.

[USGS] U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf

[USGS] U.S. Geological Survey. 2016. Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07. Available from: https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf.

Wagner RJ, Boulger RW, Jr, Oblinger CJ, Smith BA. 2006. Guidelines and standard procedures for continuous waterquality monitors: station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1-D3. Available from: https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.